Review of “Active Learning by Feature Mixing”

Zhuoyang Zou

- **Problem/Motivation**
  The paper aims to improve active learning (AL) methods, particularly for high-dimensional data (e.g., images, videos) and in low-data regimes. Existing AL methods struggle with deep neural networks and high-dimensional data, especially when there are few labeled examples available. The authors seek to develop a more effective method for selecting the most valuable samples to be labeled in the iterative AL process.
- **Algorithm**
  The authors propose a novel method called ALFA-Mix (Active Learning by FeAture Mixing). Key aspects of their solution include:
  - Interpolating between representations of labeled and unlabeled instances in the feature space.
  - Examining changes in predicted labels resulting from these interpolations to identify unlabeled instances with novel features.
  - Using a closed-form solution to efficiently find the optimal interpolation parameter.
  - Creating a candidate set of unlabeled instances whose predictions change with interpolation.
  - Clustering the candidate set and selecting diverse samples for labeling.
- **Contribution**
  The authors contribute a method that leverages feature interpolation between labeled and unlabeled instances to uncover hidden traits in the data. By examining changes in predicted labels resulting from these interpolations, ALFA-Mix identifies unlabeled instances with novel features that are likely to be informative for the model. A key innovation is the development of an efficient closed-form solution for finding the optimal interpolation parameter, making the approach computationally feasible for large-scale applications. The authors also present the first investigation of active learning with vision transformers for video classification, demonstrating the method's versatility across different data types and model architectures.
- **Downsides**
  The effectiveness of ALFA-Mix could be heavily dependent on the quality of the learned feature representations, potentially varying in performance across different models and datasets. Additionally, while the paper demonstrates improvements in computational efficiency compared to some baselines, the method's scalability to extremely large datasets or complex models may still present challenges. The approach may also be sensitive to the choice of hyperparameters, particularly the interpolation parameter ε, which could require careful tuning for optimal performance in different scenarios. Lastly, the evaluation of ALFA-Mix, like many active learning methods, is conducted in simulated environments, which may not fully capture the complexities and challenges of real-world active learning applications.

Review of "Active Learning For Open-set Annotation"

Zhuoyang Zou

- **Problem/Motivation**
  The paper aims to solve the open-set annotation problem in active learning. In real-world scenarios, unlabeled data often contains many examples from unknown classes not present in the labeled set. Traditional active learning methods fail in this setting as they tend to select these unknown class examples, wasting the annotation budget on irrelevant data. The goal is to develop an active learning approach that can effectively identify and select examples from known classes in an open-set environment.
- **Algorithm**
  The authors propose a new active learning framework called LfOSA (Learning from Open-Set Annotation). Key aspects of the solution include:
  - Decoupling detection and classification by using two separate networks.
  - Training a detector to distinguish between known and unknown classes using both labeled and invalid (unknown class) examples.
  - Using a low-temperature cross-entropy loss to enhance the distinguishability of the detector.
  - Modeling the per-example max activation value (MAV) distribution with a Gaussian Mixture Model to estimate the probability of an example belonging to a known class.
  - Actively selecting examples with the highest probability of being from known classes for annotation.
  - Updating the classifier with newly labeled examples from known classes.
- **Contribution**
  The paper makes several key contributions to active learning research. It is the first to formalize and address the OSA problem, introducing a novel framework called LfOSA. This approach decouples detection and classification tasks, using separate networks to effectively identify known classes in an open-set environment. The framework employs low-temperature cross-entropy loss and exploits both known and unknown class supervision to enhance detector performance. Additionally, it utilizes Gaussian Mixture Models to estimate per-example max activation value distribution for effective sample selection. These innovations allow LfOSA to maintain high recall in identifying examples from known classes within a large pool of unlabeled data containing unknown classes, addressing a significant challenge in real-world active learning scenarios
- **Downsides**
  The increased computational complexity due to training two separate networks and using GMMs for sample selection may limit its applicability in resource-constrained environments. The method's performance could potentially degrade in scenarios where unknown classes closely resemble known classes, leading to misclassifications. As the ratio of unknown to known classes increases significantly, the approach's effectiveness might diminish. The paper also assumes that unknown classes can be effectively modeled as a single "unknown" category, which may not hold true in all real-world applications. Lastly, while the evaluation focuses on image classification tasks, it remains unclear how well the method would generalize to other domains or more complex tasks such as object detection or segmentation. These limitations suggest that while the paper presents a promising approach, further research is needed to address these challenges and explore the method's broader applicability.

Review of “Active Learning Strategies for Weakly-Supervised Object Detection”

Zhuoyang Zou

- **Problem/Motivation**
  The paper aims to narrow the performance gap between weakly-supervised object detectors (trained only on image-level labels) and fully-supervised detectors (trained on bounding box annotations). Weakly-supervised detectors are more affordable in terms of annotation costs but perform significantly worse than fully-supervised ones.
- **Algorithm**
  The authors propose a method that combines weakly-supervised learning with active learning. They start with a base weakly-supervised detector trained on image-level labels, then iteratively select a small number of images for full bounding box annotation. These selected images are used to fine-tune the detector. The key innovation is their novel active learning strategy called "Box-in-Box" (BiB), which is designed to address common failure modes of weakly-supervised detectors.
- **Contribution**
  - A new approach combining weakly-supervised and active learning for object detection, starting without any fully-annotated data.
  - The BiB active learning strategy, tailored to address limitations of weakly-supervised detectors.
  - A difficulty-aware proposal sampling technique for more effective training.
  - Extensive experiments on VOC07 and COCO datasets, showing significant performance improvements over existing methods with very few fully-annotated images.
- **Downsides**
  It still requires manual bounding box annotations for a subset of images, which can be time-consuming and costly, even if fewer than fully-supervised approaches. Despite the significant improvements, the performance does not fully match that of fully-supervised detectors, leaving room for further enhancement. The approach may be computationally intensive due to its iterative nature, requiring multiple rounds of model retraining. Additionally, the effectiveness of the BiB strategy could vary depending on the initial weakly-supervised detector's performance. The paper focuses on VOC07 and COCO datasets, leaving questions about its applicability to more diverse or domain-specific datasets unanswered. Finally, the method's scalability to very large datasets or suitability for real-time applications may be limited due to the iterative selection and annotation process, potentially restricting its use in certain practical scenarios.